Remarks

Status of the Subject Application

Claims 1, 2, 4-6, 8-11, 13-24, 26, 27, 30-47, 49-55, and 57-61 are pending in the Subject Application. Claims 63 and 65 stand rejected under 35 U.S.C. §101 as being directed to non-statutory subject matter. Claims 1, 2, 4-6, 8-11, 13, 14, 17 and 30 stand rejected as being indefinite. Claims 1, 2, 4-6, 8-11, 13-24, 26, 27, 30-47, 49-55, and 57-61 stand rejected under 35 U.S.C. §103 as being unpatentable over United States Patent Application Publication No. 2002/0062472 to Bolle et al. (hereinafter "Bolle") in view of United States Patent No. 6,996,275 to Edanami (hereinafter "Edanami").

Claim Amendments

The claims have been amended herein to more clearly distinguish between the "low-level feature extraction" at the camera side and "high-level processing" at the server side.

Claim Rejections Under 35 U.S.C. § 101

Claims 63 and 65 have been amended in accordance with the Examiner's suggestion to overcome the rejection under 35 U.S.C. § 101.

Claim Rejections Under 35 U.S.C. § 112

With regard to the recitation of "the feature stream" in line 22 of claim 1, Applicants point out that "a feature stream" is recited in line 15 of that claim, providing appropriate antecedent basis for the recitation of "the feature stream" in line 22.

Applicants submit that claim 17 does not recite "the feature stream." However claim 15, from which claim 17 (and claim 27, which does recite "the feature stream") depends does recite "a feature stream" in line 14.

Claim Rejections Under 35 U.S.C. § 103

The Examiner rejects Claims 1, 2, 4-6, 8-11, 13-24, 26, 27, 30-47, 49-55 and 57-65 as being unpatentable over US 2002/0062482 to Bolle in view of US 6,996,275 to Edanami (newly cited reference).

In this connection, we note:

The present invention provides a method and system for performing event detection and object tracking in an image stream, utilizing an image acquisition device, which is installed in a field where the event is to be detected, and a server utility, both connected to a data network. The technique of the invention utilizes a filtering mechanism including "pre-processing" or "low-level" processing of the image data at the image acquisition device in the field, and "high-level" or "final processing" at the server. As specifically indicated in the present application, such a filtering mechanism requires minimal processing and bandwidth resources, so this technique can be concurrently applied to a large number of image streams.

The basic claims of the present application recite the low-level processing at the image acquisition device and the high-level processing at the server. The low-level processing or "low-level feature extraction" is a type of processing of an image stream to produce therefrom a feature stream upon identifying that a number and type of the features in the image stream exceed a corresponding threshold, while the high-level processing is applied to the feature stream to perform the event detection and obtain indication about the event. It is clear from the description that the "image stream" and the "feature stream" are different types of data, the features stream being selectively derived from processing the image stream. It is clear from the present application that the feature extraction from the image stream is based on threshold related data received from the server so as to enable the server to perform actual event detection from the feature stream only without an image stream at all. The image stream might be

optionally transferred to the server in addition to the feature stream upon request from the server.

Thus the invention provides distributed event detection procedure including feature extraction and creation of a feature stream at the image acquisition device and processing of the feature stream and event detection from said feature stream at the server. These features of the invention are clearly described in the present application. Indeed, see for example:

- Par. 0018: "A set of image acquisition devices is installed in field, such that each device comprises a local programmable processor for converting the acquired image stream, that consists of one or more images, to a digital format, and a local encoder, for generating features from the mage stream. The features are parameters that are related to attributes of objects in the image stream. **Each device transmits a feature stream, whenever the number and type of features exceed a corresponding threshold.** Each image acquisition device is connected to a data network through a corresponding data communication channel. An image processing server connected to the data network determines the threshold and processes the feature stream. Whenever the server receives features from a local encoder through its corresponding data communication channel and the data network, **the server obtains indications regarding events in the image streams by processing the feature stream** and transmitting the indications to an operator."

- Par. 0019: "...The local encoder may be a composite encoder, which is a local encoder that further comprises circuitry for compressing the image stream. The composite encoder may operate in a first mode, during which it generates and transmits the

features to the server, and in a second mode, during which it transmits to the server, in addition to the features, at least a portion of the image stream in a desired compression level, according to commands sent from the server. Preferably, each composite encoder is controlled by a command sent from the server, to operate in its first mode. As long as the server receives features from a composite encoder, that composite encoder is controlled by a command sent from the server, to operate in its second mode. The server obtains indications regarding events in the image streams by processing the feature stream, and transmitting the indications and/or their corresponding image streams to an operator."

- Par. 0047: ".....MCIP is based on the distribution of image processing algorithms between low-level feature extraction, which is performed by the encoders which are located in field (i.e., in the vicinity of a camera), and high-level processing applications, which are performed by a remote central server that collects and analyzes these features."

- Par. 0051: "...The features may also be generated by a specific feature extraction algorithm (such as any motion vector generating algorithm) that is not related to the video compression algorithm....."

Bolle describes a system for camera-and-operator communication. Bolle is a typical prior art technique where event detection is carried out at the camera side. Bolle is silent about any low-level processing aimed at feature extraction from an image stream and generation of a feature stream at the field-agent side, as well as silent about any high-level processing of the feature stream at the office-agent side. It is clear from the description in Bolle patent that the only type

of processing carried out at the field-agent side is extraction of a portion of the scene information which is relevant to a particular task, compressing this image and transmitting the compressed data to the office unit, where the received data is decompressed and presented to operator.

In this connection, it should be noted that the present application clearly distinguishes between compression of image stream and processing of image feature for generation of a feature stream. Indeed, see for example the above mentioned par. [0019] of the present application:

> Par. 0019: "…The local encoder may be a composite encoder, which is a local encoder that further comprises circuitry for compressing the image stream. The composite encoder may operate in **a first mode, during which it generates and transmits the features to the server**, and in **a second mode, during which it transmits to the server, in addition to the features, at least a portion of the image stream in a desired compression level**, according to commands sent from the server."

With regard to elements 350 and 360 in Fig. 3 in the Bolle patent referred to by the Examiner, these elements support the argument that Bolle's technique merely deals with compression of image stream corresponding to a specific portion of the scene. Element 550 in Fig. 5 corresponds to "semantically compressed (and then decompressed) frame 550 received at the office site 100." (see par. [0076] of Bolle reference).

Thus, **the present invention differs from the Bolle reference** not only in that "Bolle does not disclose wherein the data line is a data network and wherein the high-level processing applications is performed by a server," but in addition, the technique of the present invention is essentially different from that of Bolle in the type of processing and the configuration of the image acquisition device and the server processor. In the present invention, distributed processing is used

enabling event detection at the server from a feature stream generated from image stream at the image acquisition device. In the Bolle reference, the image acquisition device selects the image of portion of a scene where the event is detected, compresses this image data and transmits it to office unit where the image data is decompressed.

As for the **Edanami reference**, it discloses an image control apparatus which includes a monitoring site side connected via network with a monitoring station side. This technique is the typical prior art technique where the entire process of event detection is carried out at a camera side (monitoring site side).

According to Edanami reference, the monitoring station side (server side) operates as follows (see for example col. 3 lines 30-40:

> "Detection dictionary-preparing means 16 measures variance values of the feature amounts of a detection event to prepare a detection dictionary containing detection parameters defined as feature amounts having small variance values. Therefore, **whenever the user designates a detection event**, the detection dictionary-preparing means 16 updates and prepares (learns) a new detection dictionary based on the feature amounts of the detection event. Event detection control means 17 calculates the distances between feature amounts and the detection dictionary to determine a detection range for detecting the detection event."

In the Edanami's system, the threshold corresponding to the detected event is determined at the monitoring station side **based on the event detected at the camera side** for the purposes of controlling and recording the event detection results. Indeed, see for example col. 5 lines 7-20:

> "…the event detection control means 17 sets a detection threshold THmd for notifying the user that an event has been detected (that a

person has been detected, in the illustrated example), and a detection threshold THms for storing a feature amount for the detection dictionary Dm(0). More specifically, feature amounts within a radius of THms around the detection dictionary Dm(0) are each considered to have a high possibility of being the feature amount of a person, and hence are stored, while feature amounts within a radius of THmd around the detection dictionary Dm(0) are each considered to allow determination that it is of a person, and hence the user is notified of images including the respective feature amounts within this range (the images are displayed on the monitor screen)."

This technique is aimed at enabling "the user to change a detection threshold if desired" (see col. 7 lines 6-8).

Hence, the Edanami reference does not teach "server event detection" as stated by the Examiner. Rather, the Edanami reference describes how the different events detected by the camera or cameras can be classified and displayed allowing the user to operate the thresholds to be used for notifying the user about the detected event.

Thus, in view of the above, it is clear that a combination of Edanami teachings and Bolle technique *as is* would not result in the system of the invention, but significant modification would be required. Accordingly, the invention cannot be learned from the combination of these references.

Moreover, a simple test can show the difference between the technique of the invention and that resulting from combining the two references: if one would compare an effect of removal of a compressed video stream from data generated at the image acquisition device of the invention (in cases the encoder operates in its second mode and transmits such video stream in addition to the features stream) and a similar effect applied to data generated at the camera side in a "combined" system of Bolle and Edanami, then with the technique of the

invention the remaining data, after removal of the compressed video stream still provides the desired functionality of event detection at the server, whereas the "combined" system would fall apart and provide no value at all.

Accordingly, Applicants submit that the combination of Bolle and Edanami does not disclose the invention claimed in the Subject Application.

Conclusion

Applicants respectfully submit that claims 1, 2, 4-6, 8-11, 13-24, 26, 27, 30-47, 49-55, and 57-61 are in condition for allowance. Accordingly, reconsideration of the present rejections and passage to allowance of claims 1, 2, 4-6, 8-11, 13-24, 26, 27, 30-47, 49-55, and 57-61 at an early date are earnestly solicited.

If the Examiner is of the opinion that the instant application is in condition for disposition other than allowance, the Examiner is respectfully requested to contact Applicant's Attorney at the telephone number listed below. If any anticipation or obviousness rejections are contemplated by the Examiner, applicants wish to schedule an in-person interview for consideration of the grounds for such rejection at that time.

Respectfully Submitted

/Richard W. James/
Richard W. James
Registration No. 43,690
Attorney for Applicants

Spilman Thomas & Battle
One Oxford Centre, Suite 3440
301 Grant Street
Pittsburgh, Pa 15219
T (412) 325-3309
F (412) 325-3324
rjames@spilmanlaw.com